Proc. Natl. Acad. Sci. USA
Vol. 91, pp. 10834–10838, November 1994
Biochemistry

DNA length, bending, and twisting constraints on IS*50* transposition

IGOR YU GORYSHIN*†, YURI V. KIL†, AND WILLIAM S. REZNIKOFF*‡

*Department of Biochemistry, College of Agricultural and Life Sciences, University of Wisconsin, Madison, WI 53706-1569; and †Molecular Genetics Laboratory, St. Petersburg Nuclear Physics Institute, 188350 Gatchina, Leningrad District, Russia

Communicated by Carol A. Gross, July 29, 1994

ABSTRACT **Transposition is a multistep process in which a transposable element DNA sequence moves from its original genetic location to a new site. Early steps in this process include the formation of a transposition complex in which the end sequences of the transposable element are brought together in a structurally precise fashion through the action of the element-encoded transposase protein and the cleavage of the element free from the adjoining DNA. If transposition complex formation must precede DNA cleavage (or nicking), then changing the length of the donor DNA between closely spaced ends should have dramatic effects on the frequency of the transposition. This question has been examined by studying the effects of altering donor DNA length on IS*50* transposition. Donor DNA ≤ 64 bp severely impaired transposition. Donor DNA ≥ 200 bp demonstrated high transposition frequencies with only modest length dependencies. Constructs with donor DNA lengths between 66 and 174 bp demonstrated a dramatic periodic effect on transposition (periodicity ≈ 10.5 bp).**

Transposable elements are mobile DNA insertion sequences (ISs). They are believed to exist in the genomes of all organisms because they have been found in each type of organism that has been studied. Transposable elements can cause various types of genome rearrangements in addition to insertions such as deletions, inversions, and chromosome fusions and have presumably played an important role in genome evolution. For these reasons, it is of considerable interest to study the molecular details of their mechanism of transposition.

Each type of transposable element is characterized by two principal elements: specific short DNA sequences that define the ends of the element and an element-encoded transposase (Tnp) protein that catalyzes the multiple steps in the transposition process. Host proteins also function in transposition. Studies of various transposable elements have indicated that transposition involves the following general steps (1–4). The Tnp binds in a sequence-specific fashion to the end DNA sequences. The ends of the element are brought together to form a transposition complex. The Tnp either cleaves or nicks the DNA adjacent to the transposable element end sequences generating free 3′ OH ends. The 3′ OH ends perform nucleophilic attacks on staggered phosphodiester bonds at the target DNA sequence. The subsequent steps leading to an inserted transposable element follow one of two general pathways: a conservative "cut-and-paste" mechanism or a replicative mechanism depending upon the particular class of element being studied. The studies in this communication were directed at determining *in vivo* whether transposition complex formation precedes DNA cleavage/nicking for the bacterial transposable element IS*50*.

IS*50* is a bacterial transposable element that is a constituent of the compound transposon Tn*5* (5, 6). IS*50* is bounded in an inverted fashion by two 19-bp sequences that are identical at 13 positions, the outside end (OE) and the inside end (IE) (7, 8). The element encodes a 476-amino acid, cis active Tnp and a related inhibitory protein (Inh) that lacks the N-terminal 55 amino acids of Tnp (9–12). IS*50* is thought to transpose through a conservative cut-and-paste mechanism in which a DNA cleavage adjacent to the end sequences occurs (13–15), although this proposed mechanism has not been directly shown to occur for IS*50*.

We have studied the order of the steps in transposition through the following approach. IS*50* constructs were generated in which the flanking host DNA (or donor DNA) between the two ends of IS*50* was varied (see Fig. 1). Assuming that IS*50* transposes via a conservative mechanism in which the end sequences are cleaved free of the donor DNA, if Tnp must bind to both ends and form a transposition complex prior to cleavage of the DNA, there should be severe constraints on transposition for short donor DNA lengths. For very short donor DNA lengths, these constraints might result from steric inhibition of the two close protein–DNA binding reactions. For slightly longer donor DNA lengths, these constraints would presumably result from the required looping necessary to bring the two ends together (see Fig. 1). If, on the other hand, the IS*50* end sequence–donor DNA boundary was merely nicked during the early steps in transposition, we would make another set of predictions. Formation of the transposition complex prior to DNA nicking would likely result in helical orientation constraints dictated by the requirements to avoid steric hindrance and to have the Tnp bound ends in the correct relative orientation. Finally, for long relaxed donor DNA lengths (as would happen if nicking preceded transposition complex formation), one would expect a length-related "concentration"-dependent reduction in transposition frequencies since the two ends would need to find each other through a three-dimensional diffusion process. Failure to find such a length dependency suggests that transposition complexes form on supercoiled molecules (that are not nicked).

A preliminary report of this work has been published (16).

MATERIALS AND METHODS

Plasmid Constructs. The pIS50.11-*n* plasmids (where *n* is the spacer DNA length) contain an IS*50* construct in which the Tnp gene (and the Inh gene) is programed by its normal regulatory elements and is situated in its normal location vis-à-vis the OE sequence (see Fig. 2). Between *tnp* and IE, a 2226-bp fragment from pACYC-184 encoding chloramphenicol (Cam) acetyltransferase (*cat*) and an origin of replication (*ori*) has been inserted. The construction of the pIS50.11-*n* plasmids was initiated with an IS*50* precise excisant in which 9 bp of donor DNA was located between the ends (pIS50.11-9, ref. 18). This excisant fortuitously contained a *Dra* II site at the donor DNA–OE boundary. The next step in

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviations: OE, outside end; IE, inside end; Tnp, transposase; Inh, inhibitor; Kan, kanamycin; Gen, gentamycin; Nal, nalidixic acid; Cam, chloramphenicol; IS, insertion sequence.

‡To whom reprint requests should be addressed.

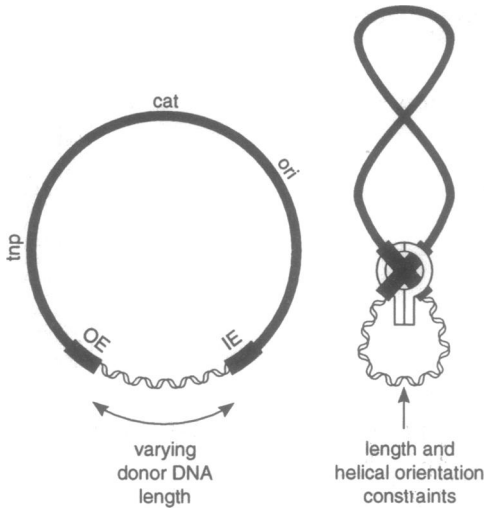

FIG. 1. Donor DNA looping during transposition complex formation. The plasmid constructs that we have used are designed so that all essential functions (including the origin of replication, the Tnp gene, and the chloramphenicol acetyltransferase gene) are encoded within IS*50* (see Fig. 2). This IS*50* DNA is sufficiently long (≈4 kbp) to not impose any limits on flexibility. The donor DNA, on the other hand, can be made quite short so that detailed length dependencies can be examined. If cleavage of the DNA at the ends of IS*50* precedes transposition complex formation, no donor DNA length constraints should be observed.

the construction involved cloning a kanamycin-resistance (Kan^r) cassette surrounded by two *Eco*RI–*Bam*HI–*Sal* I-*Pst* I polylinkers into the *Dra* II site, thus creating pIS50.11-1298. Removal of the kan^r cassette by *Sal* I digestion followed by ligation gave the 46-bp construct. The same operation using *Bam*HI generated the 34-bp construct. Other plasmids in this series were generated by cloning random *Sau*3A fragments from pTZ18R (19) into the 34-bp construct unique *Bam*HI site. One such insert resulted in the formation of the 80-bp construct pIS50.11-80.

The pFMA50-*n* plasmid series (see Fig. 2) was derived from pIS50.11-80 by inserting a *Sau*3A 740-bp fragment containing the f1 *ori* into a *Bcl* I site, thus creating a phagemid. The MA56 mutation was incorporated into *tnp* by site-directed mutagenesis. MA56 allows an ≈10-fold increase in the sensitivity of transposition assays because it eliminates the synthesis of Inh (17). The donor DNA sequences for the pFMA50-9 through -109 are shown in Fig. 3 and the generation of the entire series is described in the legend to Fig. 3.

Transposition Assays. The transposition frequencies were determined by a mating out assay similar to that described earlier (17). Donor cells {RZ212 [Δ(*lac-proA,B*), *ara, str, recA56, srl, thi*/pOX38-Gen] (17)} containing the relevant pIS50.11-*n* or pFMA50-*n* plasmid were grown to an OD_{600} = 0.8 and mixed at a ratio of 1:3 with recipient cells {14R525 [an F$^-$ nalidixic acid-resistant (Nal^r) prototroph (20)]} at the same OD. After 3 hr of gentle aeration, aliquots of the mixture were plated on LB agar (21) containing chloramphenicol (Cam; 20 μg/ml) plus Nal (20 μg/ml) and LB agar containing gentamycin (Gen; 5 μg/ml) plus Nal (20 μg/ml), and the ratio of colonies growing on the two media was taken as the transposition frequency. The transposition tests were performed 5–20 times for each construct.

RESULTS

To examine whether Tnp binding to both end sequences and transposition complex formation (in which the two ends of IS*50* are brought together) precede a DNA cleavage or nicking event in the IS*50* transposition process, we examined the transposition frequency for IS*50*-containing plasmids in which the donor DNA length varied for different constructs. We reasoned that if Tnp binding to both ends and transposition complex formation occurred first, then the transposition frequency should be quite sensitive to the donor DNA length for short donor DNAs either because DNA looping is required (Fig. 1) or because of steric interference of one end bound Tnp on the binding of the second Tnp. To perform these analyses, two sets of plasmids were constructed in which the donor DNA (but not IS*50* DNA) could be varied at will. This was accomplished by using starting constructs in which the origin of replication and other required functions were all encoded within IS*50*, allowing us total flexibility in manipulating the donor DNA (see Fig. 2 for a description of these constructs).

The overall general effects of donor DNA length on IS*50* transposition were studied using the pIS50.11-*n* plasmid series. The results are shown in Fig. 4. It is obvious that there

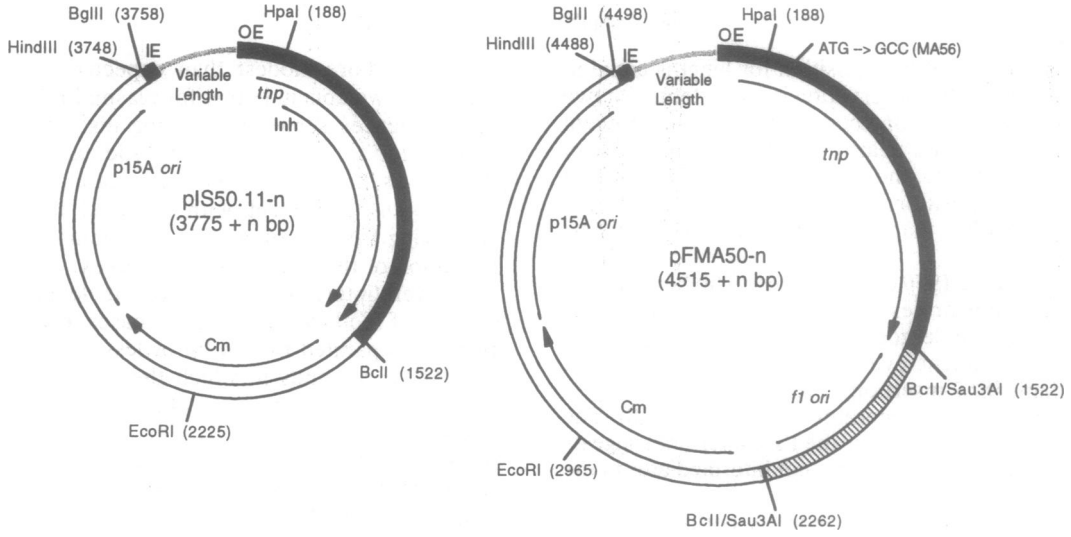

FIG. 2. Structure of plasmids pIS50.11-*n* and pFMA50-*n*. The pIS50.11-*n* plasmids (where *n* is the spacer DNA length) contain an IS*50* construct in which the Tnp gene (and the Inh gene, which is coincidental with *tnp* except for the N-terminal 55 codons) is programed by its normal regulatory elements and is situated in its normal location vis-à-vis the OE sequence. Between *tnp* and the IE is a gene encoding Cam acetyltransferase (*cat*) and an origin of replication (*ori*). The pFMA50-*n* plasmid series was derived from pIS50.11-80 by inserting a *Sau*3A 740-bp fragment containing the f1 *ori* into a *Bcl* I site, thus creating a phagemid. The MA56 mutation was incorporated into *tnp* by site-directed mutagenesis. MA56 allows an ≈10-fold increase in the sensitivity of transposition assays because it eliminates the synthesis of Inh (17).

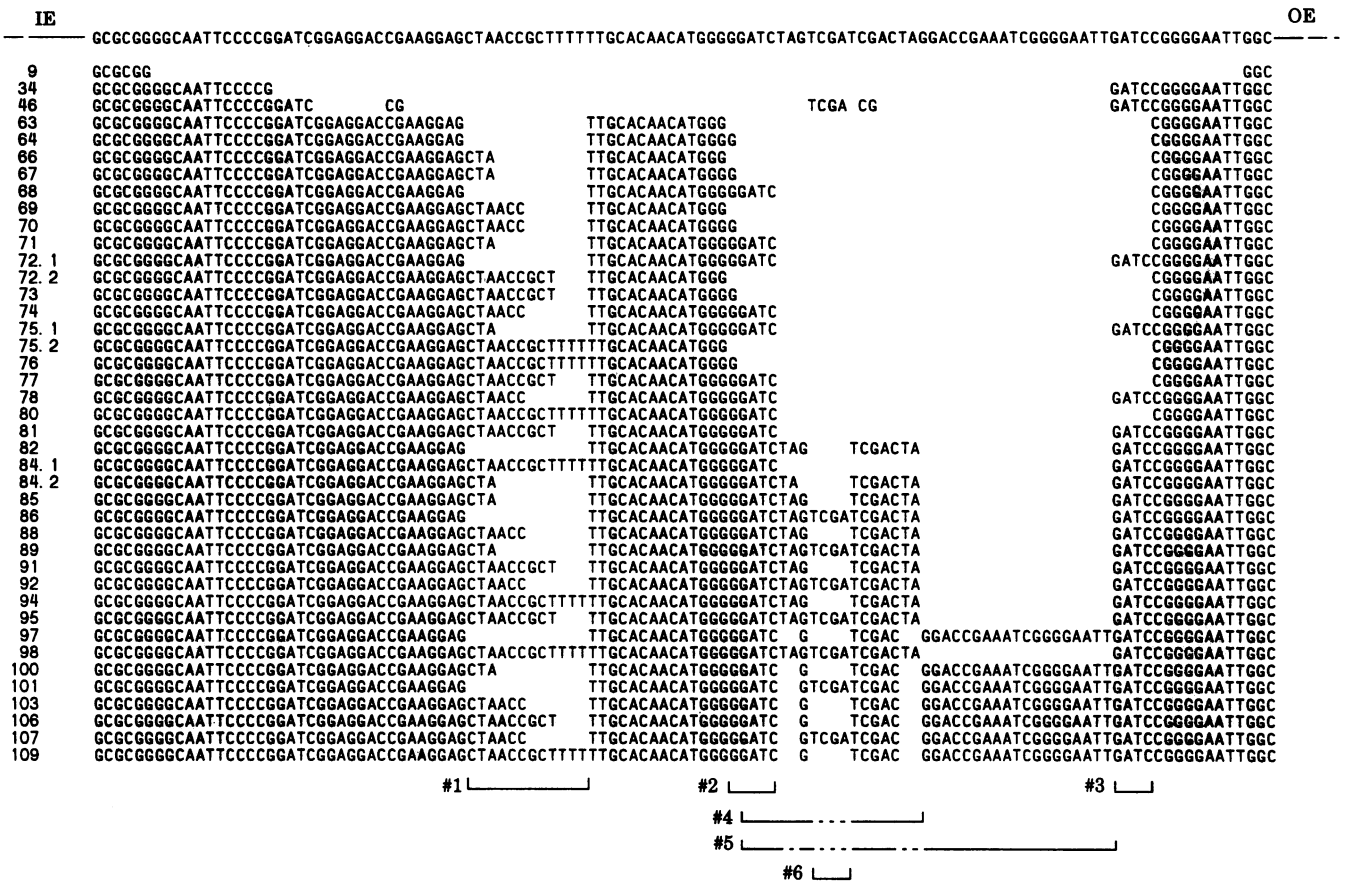

FIG. 3. DNAs between 9 and 109 bp used to examine donor DNA length dependence of IS*50* transposition. The upper line shows a consensus sequence for the variable DNA in the pFMA50-*n* plasmid series. The left column shows the exact number of base pairs between OE and IE, corresponding to the number after the hyphen in the name of the particular construct. Plasmids in the pFMA50-*n* series originated from pFMA50-80. pFMA50-9, -34, and -46 were created by exchange of *Eco*RI and *Hpa* I fragments between pFMA50-80 and corresponding pIS50.11-9, -34, and -46 plasmids (see Fig. 2). For other plasmids between pFMA50-63 and pFMA50-109, the regions in pMFA50-80 indicated below the sequence were changed as follows: #1 includes deletions of 3, 6, 9, and 12 bp generated by site-directed mutagenesis; #2 includes 4- and 5-bp deletions generated by S1 treatment and ligation after *Bam*HI digestion (a *Bam*HI site was generated at one end by insertion of the *Sau*3A fragment); #3 involved a 4-bp insert caused by filling in the *Bam*HI site prior to ligation; #4 involved insertion of a +14-bp linker into the *Bam*HI site prior to ligation; #5 involved insertion of a +29-bp linker into the *Bam*HI site prior to ligation; #6 involved a 4-bp insert caused by filling in a *Sal* I site introduced by the +14 and +29 linkers. The +29-bp linker was designed to resemble the most common DNA in the series. All donor DNAs were confirmed by sequence analysis. DNAs 72.1 and 72.2, 75.1 and 75.2, and 84.1 and 84.2 are in each case two different constructs with identical lengths. The sequences of the 112, 116, 142, 145, 148, 151, 154, 157, 160, 163, 166, 169, and 172 donor DNAs are closely related to the above series and a description is available by request.

is a dramatic inhibition of transposition for lengths equal to or shorter than 64 bp. This observation is consistent with two possibilities. Tnp bound to one end might sterically prevent Tnp binding to the second nearby end for short lengths of donor DNA. Alternatively, the DNA between the ends is not cleaved prior to transposition complex formation and it is the required DNA bending that limits transposition for short donor DNAs.

We also note that the IS*50* construct with 9 bp between the end sequences demonstrates a very low transposition frequency. When IS*50* undergoes precise excision from host DNA [a Tnp-independent event in which the 9-bp direct repeats that flank IS*50* insertions "recombine," leading to IS*50* insertion loss (22)], it generates this structure (18). The low frequency of transposition for the IS*50* excisant suggests that these circularized structures are not intermediates in the transposition process.

For sequences beyond 400 bp there is perhaps a slight decrease in the transposition frequency. If the two ends found each other through a random three-dimensional search, we would expect a length dependence of $l^{-3/2}$ (23), which clearly does not occur. Further experiments with donor DNAs of 3.8 kbp in length confirm the observation that the reduction in transposition frequencies with increasing lengths is more modest than expected (data not shown). Therefore we conclude that for plasmid DNA, the two IS*50* end sequences separated by long distances can find each other prior to transposition by sliding (snaking) of DNA through itself (a one-dimensional search) (24–26). This observation is consistent with a model in which the IS*50* containing DNA is supercoiled (contains no nicks) prior to formation of the transposition complex.

We were interested in studying the detailed length dependencies of IS*50* transposition for cases in which the OE and IE were separated by 174 bp or less. For this we used the pFMA50-*n* plasmids in which the donor DNAs are all sequence related and many of which differ by single base-pair steps (see Figs. 2 and 3). The transposition results are presented in Fig. 5. It is clear that there is a dramatic periodic relationship between donor DNA length and transposition frequency with differences of >200-fold for some 5-bp shifts. From this we can conclude that there are no DNA nicks between the two end sequences prior to the Tnp binding to both ends. If we hypothesize that we are looking at transposition complex formation, the results suggest that the architecture of the IS*50* transposition complex has a precise inflexible structure that requires a specific orientation of the

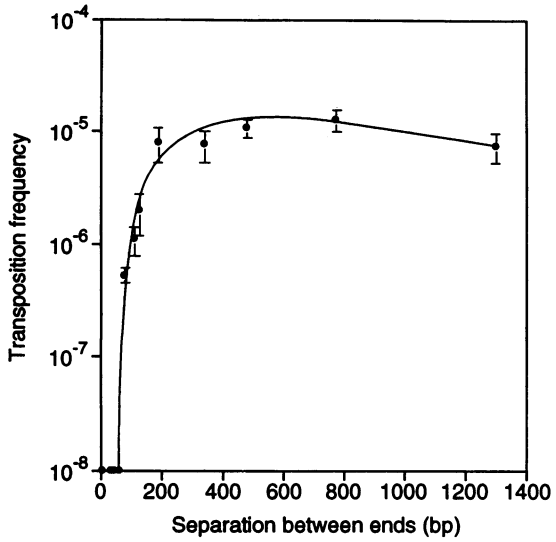

FIG. 4. Large-scale length dependence of IS*50* transposition. The indicated donor DNA lengths were 9, 34, 46, 64, 80, 112, 127, 187, 340, 480, 774, and 1298 bp. The transposition tests were performed by a mating out assay (see text).

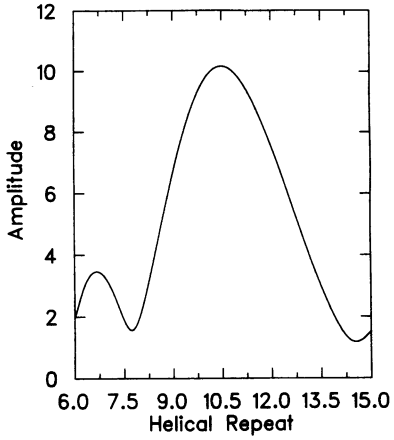

FIG. 6. Fourier analysis of transposition data for donor DNA lengths between 63 and 116 as pictured in Fig. 5. The periodicity is determined to be 10.49 bp. The calculated periodicity for the entire 63- to 172-bp region is 10.54 bp.

two end sequences relative to each other (or relative to the associated Tnp oligomer).

The periodic pattern displayed in Fig. 5 was subjected to a Fourier analysis (27). As demonstrated in Fig. 6, the periodicity in our data is 10.5 bp.

DISCUSSION

DNA transposition is a multistep process including steps in which specific transposition complex formation and DNA cleavage (or nicking) adjacent to the ends are critical events. The data presented in this communication clearly show that for IS*50* there is a dramatic inhibition of transposition when the end sequences are brought too close together (≤64 bp) and, moreover, that for somewhat longer donor DNA lengths there is a periodic relationship between DNA length and transposition. The simplest explanations for these observations are that transposition complex formation must precede DNA cleavage or nicking and that the transposition complex has a rather precise rigid architecture. The hypothesis that transposition complex formation precedes DNA cleavage is consistent with our previous observation that changing the sequence of one Tn*5* end sequence can alter the nature and frequency of adjacent deletion events associated with a second end sequence 12 kbp away (28). Thus the two end sequences must interact prior to the cleavage or nicking event associated with adjacent deletion formation.

We cannot rule out the possibility that some of the donor DNA length effects might result from Tnp bound to one end

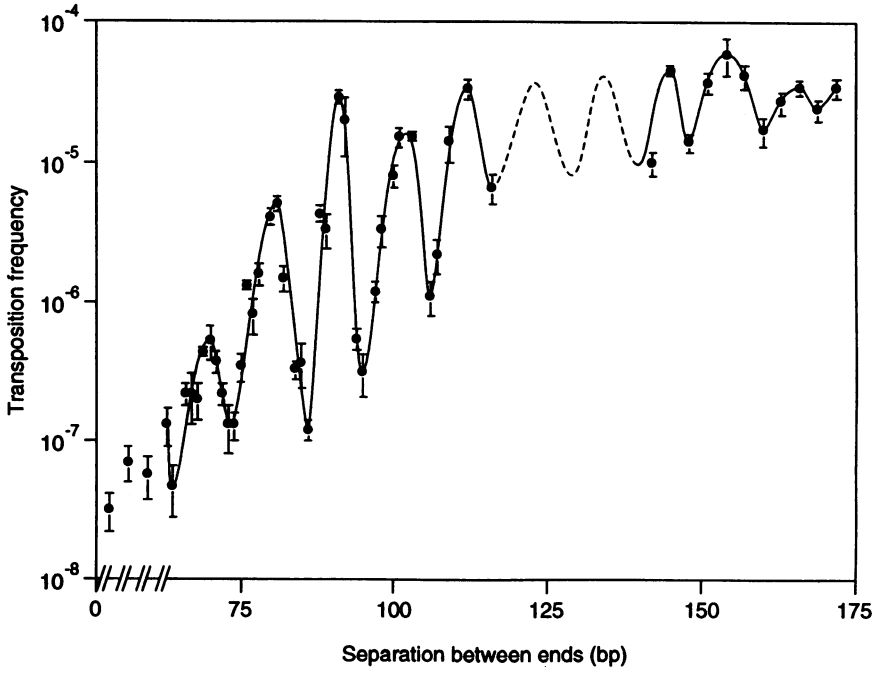

FIG. 5. Detailed analysis of IS*50* transposition frequencies for donor lengths of 9–172 bp. The indicated donor DNA length plasmids (pFMA50-80 derivatives; see Fig. 2) were constructed as described in the legend to Fig. 3. The transposition tests were performed as described in the text. Another set of spacing sequences in the range of 180–192 bp still shows periodicity with an amplitude of ≈3-fold (data not shown). The background for the transposition assays was estimated with an IS*50* construct containing only one transposable element end and was $1.2 \pm 0.6 \times 10^{-8}$.

sterically hindering Tnp binding to the other end, especially for the constructs with very short donor DNA sequences. However, the fact that the overall pattern of transposition does not increase abruptly but rather increases with molecules differing in donor DNA lengths from 64 to 91 bp and the fact that periodic changes in transposition are still observed for molecules with donor DNA sequences out to 192 bp in length argue that steric inhibition cannot explain all of the results. We favor the alternative model that DNA looping constraints affect transposition frequencies.

Lane *et al.* (29) have recently performed experiments with IS*1* in which the internal distance between the end sequences was varied in length. These investigators noticed a DNA length-dependent periodicity (10–11 bp) of Tnp-dependent SOS induction and cointegrate formation. Since these length differences were internal to the IS*1* ends, their results do not address the issue of whether IS*1* transposition complex formation precedes DNA cleavage or nicking. However, their results are compatible with our model that transposition complexes have a rather precise architecture and that the minimum required loop length may be on the order of 65 bp in the IS*1* system.

Underlying the proposed model pictured in Fig. 1 is the assumption that IS*50* transposition proceeds by a conservative cut-and-paste mechanism in which IS*50* ends are cleaved free from donor sequences. This has not been rigorously shown for IS*50*, although it is consistent with current models (13–15). An alternative model involves the generation of nicks at the transposon ends (1). Our data also suggest that transposition complex formation occurs prior to nick generation, not because of the apparent minimal length dependence (that might still exist for nicked molecules) but because of the periodicity in transposition frequencies observed in Fig. 5.

Previous investigators have found *in vivo* DNA periodicities of 11.0–11.3 bp (27, 30, 31). As indicated by the Fourier analysis of our data (Fig. 6), the periodicity of the relative transposition frequency that we observed is 10.5 bp, which is identical to the helical repeat of various linear DNAs in solution (32–34). It is possible that the short *in vivo* periodicity that we observed is due to the generation of local domains of supercoiling caused by nearby transcription (35). In particular, we note that the Tnp and Inh promoters are located near the OE, and the *ori* contains promoters near the IE. Alternatively, the short helical repeat may be related to the fact that IS*50* transposition occurs primarily off of newly replicated DNA (36). It is, however, possible that the mean periodicity of the donor DNA sequences is ≈11 but that transposition complex formation selects molecules that are transiently overwound.

A thermodynamic model developed by Law *et al.* (27) predicts an increasing phasing amplitude with a decreasing spacing distance. Our data are consistent with this model only for lengths ≥85 bp. For shorter distances the amplitude of the effect appears to become smaller. One explanation for this disagreement is that the required twisting for short distances might in part be facilitated by the necessary sharp bending for these lengths.

Our failure to detect a significant length dependency for distances greater than 400 bp suggests that the OE and IE sites can find each other by sliding of the DNA through itself in a one-dimensional search process. This conclusion is consistent with proposals that DNA supercoiling facilitates the process of bringing distant sites together (24–26) and with the hypothesis that the molecules undergoing transposition do not have nicks next to their end sequences prior to transposition complex formation.

We especially thank Dr. Vladislav A. Lanzov, Director of the Molecular Genetics Laboratory in St. Petersburg, for his support, interest, and encouragement. We also thank Paula Schlax for performing the Fourier analysis, Michael Weinreich and Debbie Czarniecki for helpful suggestions, and Michael Chandler for sharing unpublished results. This research was supported by a United States–Russia Cooperative Research Grant from the National Science Foundation (MCB 920681) and by a grant from the American Cancer Society (MV-64163). W.S.R. is the Evelyn Mercer Professor of Biochemistry and Molecular Biology.

1. Mizuuchi, K. (1992) *Annu. Rev. Biochem.* **61**, 1011–1051.
2. Bainton, R., Gamas, P. & Craig, N. L. (1991) *Cell* **65**, 805–816.
3. Bender, J. & Kleckner, N. (1986) *Cell* **45**, 801–815.
4. Berg, D. E. & Howe, M. M. (1989) *Mobile DNA* (Am. Soc. for Microbiol., Washington, DC).
5. Berg, D. E. (1989) in *Mobile DNA*, eds. Berg, D. E. & Howe, M. M. (Am. Soc. for Microbiol., Washington, DC), pp. 184–210.
6. Reznikoff, W. S. (1993) *Annu. Rev. Microbiol.* **47**, 945–963.
7. Johnson, R. C. & Reznikoff, W. S. (1983) *Nature (London)* **304**, 280–282.
8. Sasakawa, C., Carle, G. F. & Berg, D. E. (1983) *Proc. Natl. Acad. Sci. USA* **80**, 7293–7297.
9. Isberg, R. R., Lazaar, A. L. & Syvanen, M. (1982) *Cell* **30**, 883–892.
10. Johnson, R. C. & Reznikoff, W. S. (1984) *J. Mol. Biol.* **177**, 645–661.
11. Johnson, R. C., Yin, J. C. & Reznikoff, W. S. (1982) *Cell* **30**, 873–882.
12. Krebs, M. P. & Reznikoff, W. S. (1986) *J. Mol. Biol.* **192**, 781–791.
13. Berg, D. E. (1983) *Proc. Natl. Acad. Sci. USA* **80**, 792–796.
14. Hirschel, B. J., Galas, D. J., Berg, D. E. & Chandler, M. (1982) *J. Mol. Biol.* **159**, 557–580.
15. Hirschel, B. J., Galas, D. J. & Chandler, M. (1982) *Proc. Natl. Acad. Sci. USA* **79**, 4530–4534.
16. Goryshin, I. Y. & Reznikoff, W. S. (1994) *J. Cell. Biol. Suppl.* **18B**, F139 (abstr.).
17. Wiegand, T. W. & Reznikoff, W. S. (1992) *J. Bacteriol.* **174**, 1229–1239.
18. Kil, Y. V., Goryshin, I. Y. & Lanzov, V. A. (1994) *Mol. Biol. (Moscow)* **28**, 563–573.
19. Mead, D. A., Szczesna-Skoruea, E. & Kemper, B. (1986) *Protein Eng.* **1**, 67–74.
20. Wiater, L. A. & Grindley, N. D. F. (1990) *J. Bacteriol.* **172**, 4951–4958.
21. Miller, J. H. (1992) *A Short Course in Bacterial Genetics* (Cold Spring Harbor Lab. Press, Plainview, NY).
22. Egner, C. & Berg, D. E. (1981) *Proc. Natl. Acad. Sci. USA* **78**, 459–463.
23. Jacobson, H. & Stockmayer, W. H. (1950) *J. Chem. Phys.* **18**, 1600–1606.
24. Vologodskii, A. V., Levene, S. D., Klenin, K. V., Frank-Kamenetskii, M. D. & Cozzarelli, N. R. (1992) *J. Mol. Biol.* **227**, 1224–1244.
25. Kanaar, R. & Cozzarelli, N. R. (1992) *Curr. Opin. Struct. Biol.* **2**, 369–379.
26. Adams, D. E., Bliska, J. B. & Cozzarelli, N. R. (1992) *J. Mol. Biol.* **226**, 661–673.
27. Law, S. M., Bellomy, G. R., Schlax, P. J. & Record, M. T., Jr. (1993) *J. Mol. Biol.* **230**, 161–173.
28. Jilk, R. A., Makris, J. C., Borchardt, L. & Reznikoff, W. S. (1993) *J. Bacteriol.* **175**, 1264–1271.
29. Lane, D., Cavaillé, J. & Chandler, M. (1994) *J. Mol. Biol.*, in press.
30. Lee, D. H. & Schleif, R. F. (1989) *Proc. Natl. Acad. Sci. USA* **86**, 476–480.
31. Haykinson, M. J. & Johnson, R. C. (1993) *EMBO J.* **12**, 2503–2512.
32. Rhodes, D. & Klug, A. (1980) *Nature (London)* **286**, 573–578.
33. Tullius, T. D. & Dombrowski, B. A. (1985) *Science* **230**, 679–681.
34. Niederweis, M., Lederer, T. & Hillen, W. (1992) *J. Mol. Biol.* **228**, 322–326.
35. Liu, L. F. & Wang, J. C. (1987) *Proc. Natl. Acad. Sci. USA* **84**, 7024–7027.
36. Yin, J. C.-P., Krebs, M. P. & Reznikoff, W. S. (1988) *J. Mol. Biol.* **199**, 35–46.